Finding Changing Crime Regions

Use of High Dimensional Geographic Feature Space and Classification Trees

Michael D. Porter and Donald E. Brown

{mdp2u,brown}@virginia.edu

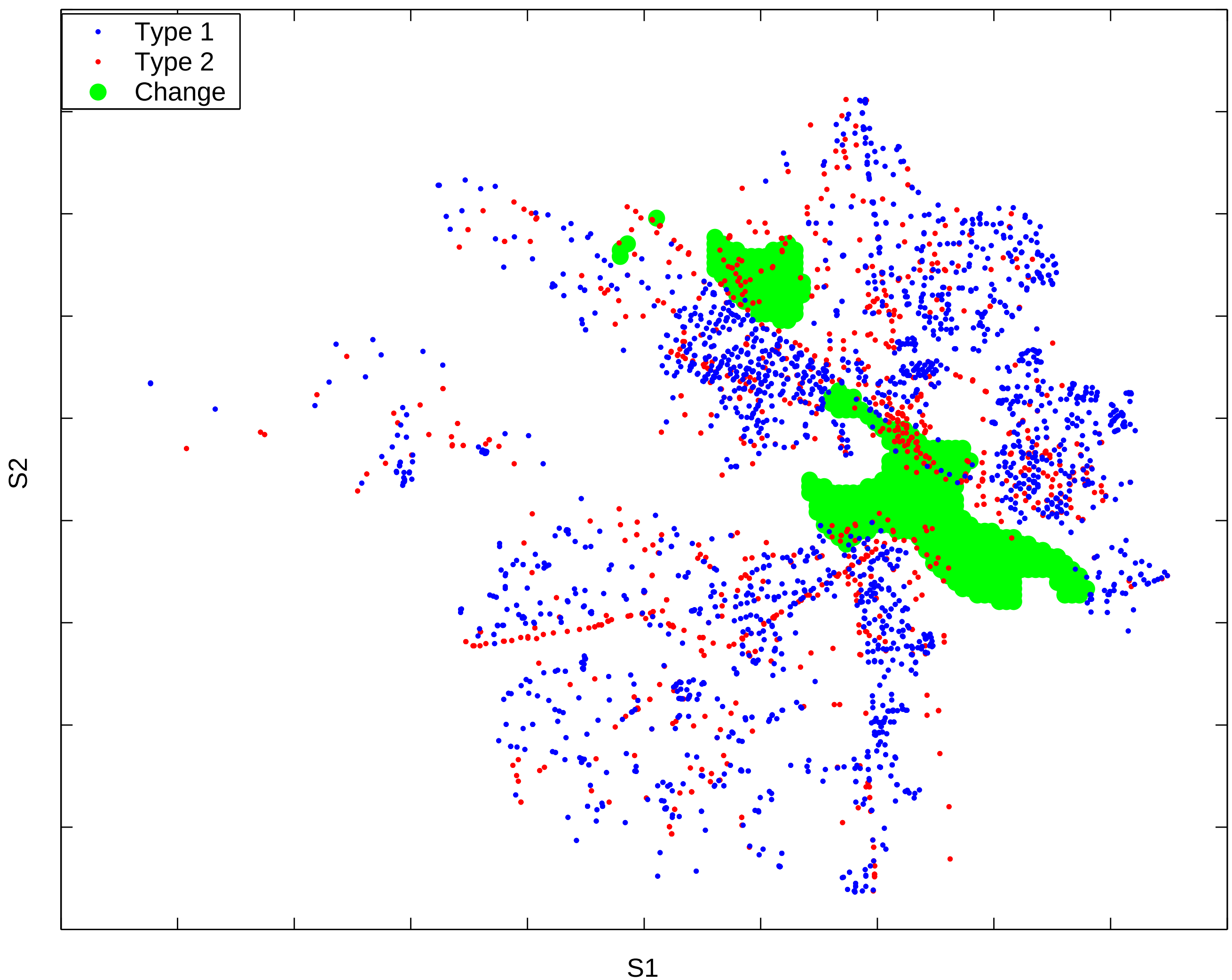
Department of Systems and Information Engineering

University of Virginia

Charlottesville, Virginia 22904

UNIVERSITY *of* VIRGINIA

Example Change Region



Agenda

- Description of Changes to be Detected
- Approach to Criminal Change Detection
- Problem Formulation - Point Process and Hypothesis Test
- Methodology - Test Statistic
- Significance - Monte Carlo
- Examples
- Discussion

Problem Statement

Statistically Detect Changes in Criminal Processes

- Treat as intelligent site selection problem
  - Detect changes in intensity or activity level of process
  - Detect changes in behavior/preferences of criminals
- Detect local regions of change
- Examples
  - Changes in the criminal process between two time periods
  - Differences between two types of crimes
  - Differences between case-control data sets

Intelligent Site Selection

- Social Ecological theories seek to describe the motivations and acts of crime based on the general features of one's social environment
  - Routine Activities Theory (Cohen & Felson 1979)
  - Rational Choice Theory (Cornish & Clarke 1986)
- We wish to capture the environmental factors which are influencing criminal actions
  - Number and Location of events
- Detect changes in these factors in addition to spatial regions

Additional Features

- Considering additional environment features extends the traditional spatial methods
- Not using all your information with spatial analyses alone
- Examples of additional features:
  - Census Features
  - GIS Data (Distance to landmarks or structures)
  - Indicators (Neighborhood Watch, street lights, wooded)

Problems

- Cannot capture all features that are considered by criminals
- Different types of decision makers (criminals) consider different features
  - Not all individual features considered by each criminal
- Very high dimensions
- Feature space is not uniformly distributed in study region
- Traditional modeling efforts will fail

Point Process

- A stochastic model governing the location (and number) of events in some set [Cressie, 1993]
- A point process P={events: event is element of set $\mathcal{X}$ }, where $\mathcal{X}$ is the set over which the point process is defined
- Specified by

$$\mu(B) = \int_B \lambda(\mathbf{s})\, d\mathbf{s}$$

  where μ - mean measure and $\lambda(\mathbf{s})$ - intensity function

- $\Pr(N(B) = n) = \frac{\exp\{-\mu(B)\}[\mu(B)]^n}{n!}$ (Poisson)

Marked Point Process

Let the criminal process be represented by $P = \{(s_i, k_i)\}$, a marked point process on space $\mathcal{X} = A$ x K, $A \subset \mathbb{R}^2, K = \{1, 2\}$.

Assuming the ground process is a nonhomogeneous Poisson spatial point process and the marks are independent of each other, P can be specified by:

$$\{\lambda(\mathbf{s}, k) = \lambda_g(\mathbf{s}) \cdot f(k|\mathbf{s}) \ : \ \mathbf{s} \in A, k \in \{1, 2\}\}$$

An observation is:

$$\Omega = [(s_1, k_1), (s_2, k_2), \ldots, (s_N, k_N), N_1 = n_1, N_2 = n_2]$$

Feature Space

The *feature space* G, defines the additional geographic and socio-economic information relating to the locations in the study region.

Let the values of an event's location in feature space be designated by:

$$\mathbf{g}(\mathbf{s}) = (g_1, g_2, \ldots, g_p, g_{p+1}, \ldots, g_{p+q}) \in G$$

The value $\mathrm{g}(\mathrm{s})$ is assumed to be a known function of location, so given any $\{\mathrm{s}\}$,the value $\mathrm{g}(\mathrm{s})$ can be determined.

Feature Space

From the previous discussion, we said the locations and number of crimes are influenced by the values of the region's feature set. We are in fact implying the the values of our intensity function are dependent upon these feature values.

$$\lambda(\mathbf{s}, k) = \lambda_g(\mathbf{s}) \cdot f(k \mid \mathbf{s})$$

$$\Leftrightarrow$$

$$\lambda(\mathbf{s}, k; g(\mathbf{s})) = \lambda_g(\mathbf{s}; g(\mathbf{s})) \cdot f(k \mid \mathbf{s}; g(\mathbf{s}))$$

Hypothesis Test I

$$H_o : \lambda(\mathbf{s}, k = 1) = \lambda(\mathbf{s}, k = 2)$$

$$\Rightarrow \lambda_g(\mathbf{s}) \cdot f(k = 1 \mid \mathbf{s}) = \lambda_g(\mathbf{s}) \cdot f(k = 2 \mid \mathbf{s})$$

$$\Rightarrow \frac{f(k = 1 \mid \mathbf{s})}{f(k = 2 \mid \mathbf{s})} = \theta(\mathbf{s}) = 1$$

Hypothesis Test I

$$H_o : \lambda(\mathbf{s}, k = 1) = \lambda(\mathbf{s}, k = 2)$$

$$\Rightarrow \lambda_g(\mathbf{s}) \cdot f(k = 1 \mid \mathbf{s}) = \lambda_g(\mathbf{s}) \cdot f(k = 2 \mid \mathbf{s})$$

$$\Rightarrow \frac{f(k = 1 \mid \mathbf{s})}{f(k = 2 \mid \mathbf{s})} = \theta(\mathbf{s}) = 1$$

$$\boxed{\begin{aligned} H_o : \theta(\mathbf{s}) &= 1 && \forall s \in A \\ H_a : \theta(\mathbf{s}) &= \begin{cases} \delta \neq 1 & \forall s \in B \subseteq A \\ 1 & \forall s \in A \setminus B \end{cases} \end{aligned}}$$

Hypothesis Test I

$$H_o : \lambda(\mathbf{s}, k = 1) = \lambda(\mathbf{s}, k = 2)$$

$$\Rightarrow \lambda_g(\mathbf{s}) \cdot f(k = 1 \mid \mathbf{s}) = \lambda_g(\mathbf{s}) \cdot f(k = 2 \mid \mathbf{s})$$

$$\Rightarrow \frac{f(k = 1 \mid \mathbf{s})}{f(k = 2 \mid \mathbf{s})} = \theta(\mathbf{s}) = 1$$

$$\boxed{\begin{aligned} H_o : \theta(\mathbf{s}) &= 1 & \forall s \in A \\ H_a : \theta(\mathbf{s}) &= \begin{cases} \delta \neq 1 & \forall s \in B \subseteq A \\ 1 & \forall s \in A \setminus B \end{cases} \end{aligned}}$$

Since $f(k = 1 \mid \mathbf{s}) + f(k = 2 \mid \mathbf{s}) = 1$ for all $\mathbf{s}$, this implies

$$f(k = 1 \mid \mathbf{s}) = \theta(\mathbf{s})(\theta(\mathbf{s}) + 1)^{-1}$$

$$f(k = 2 \mid \mathbf{s}) = (\theta(\mathbf{s}) + 1)^{-1}$$

Statistical Test-Likelihood Ratio

$$T_I(\Omega) = \frac{\mathcal{L}(\theta_{H_o})}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)} = \prod_{i=1}^{N_{1,2}} \frac{\lambda_g(\mathbf{s}_i) \cdot f(k_i \mid \mathbf{s}_i; \theta_{H_o}) \cdot \exp\{\int_A \lambda_g(\mathbf{s}) d\mathbf{s}\}}{\lambda_g(\mathbf{s}_i) \cdot f(k_i \mid \mathbf{s}_i; \hat{\theta}_{ML}) \cdot \exp\{\int_A \lambda_g(\mathbf{s}) d\mathbf{s}\}}$$

$$= \prod_{i=1}^{N_{1,2}(B)} \left[\frac{\theta_{H_o}(\theta_{H_o}+1)^{-1}}{\hat{\theta}(\hat{\theta}+1)^{-1}} \right]^{y_i} \left[\frac{(\theta_{H_o}+1)^{-1}}{(\hat{\theta}+1)^{-1}} \right]^{(1-y_i)}$$

where $y_i = 1$ if $k_i = 1$ and $N_{1,2}(B)$=Number of Type 1 and Type 2 events in region B

$$= \left[\frac{1}{2} \right]^{N_{1,2}(B)} \left[\frac{N_{1,2}(B)}{N_1(B)} \right]^{N_1(B)} \left[\frac{N_{1,2}(B)}{N_2(B)} \right]^{N_2(B)}$$

Since $\hat{\theta}_{\text{ML}} = \frac{N_1(B)}{N_2(B)}$, and $\theta_{H_o} = 1$

- If $T_I(\Omega) \leq \gamma$, reject H_o.
- $T_I(\Omega)$ only depends on region B.

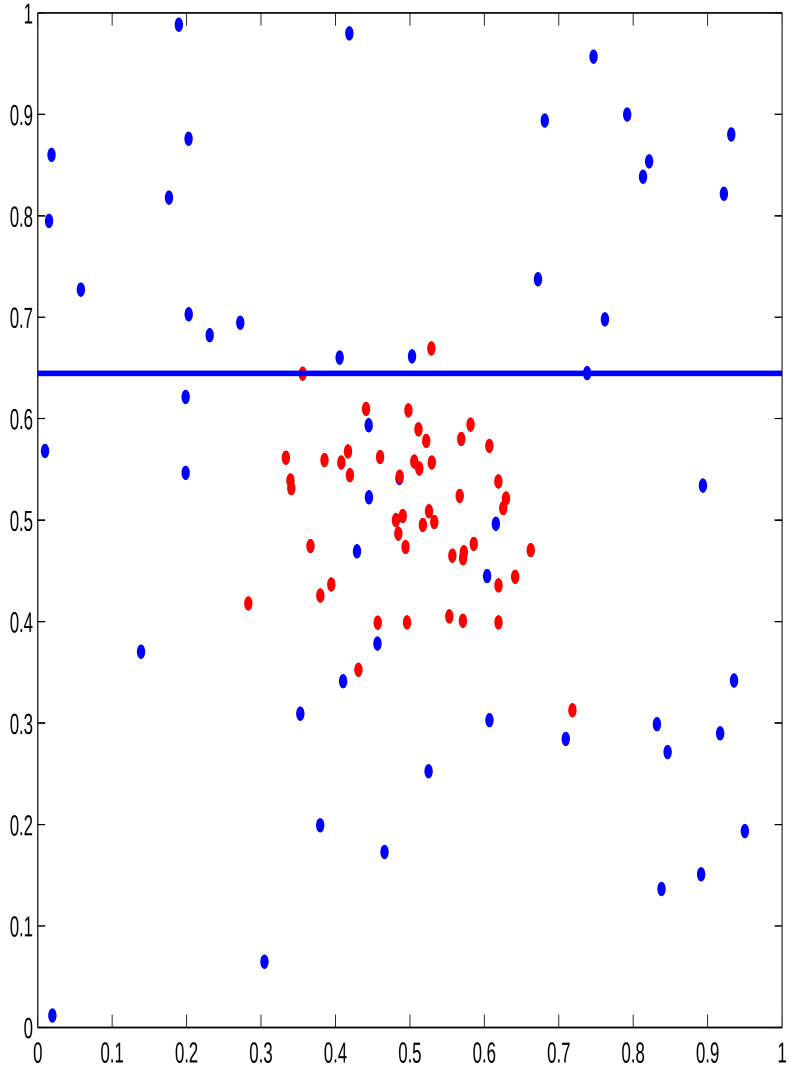
Relationship to Classification

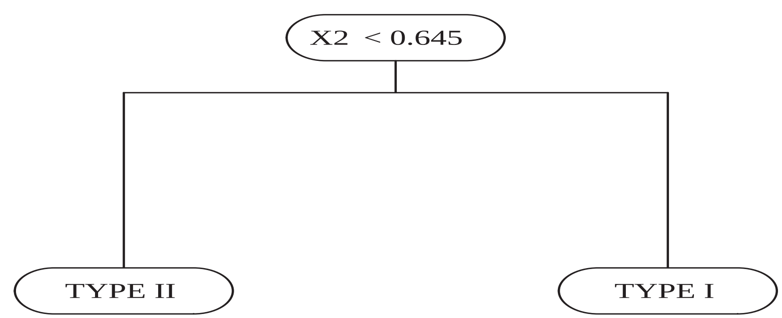
- Seeking the region B^* that minimizes $T_I(\Omega)$
- We can employ classification methods to seek out possible candidates for B^*
  - Logistic Regression
  - GAM's
  - Kernel Methods
  - Neural Nets
  - **CART (Classification Trees)**
- Then we must test for significance of B^*

Classification Trees

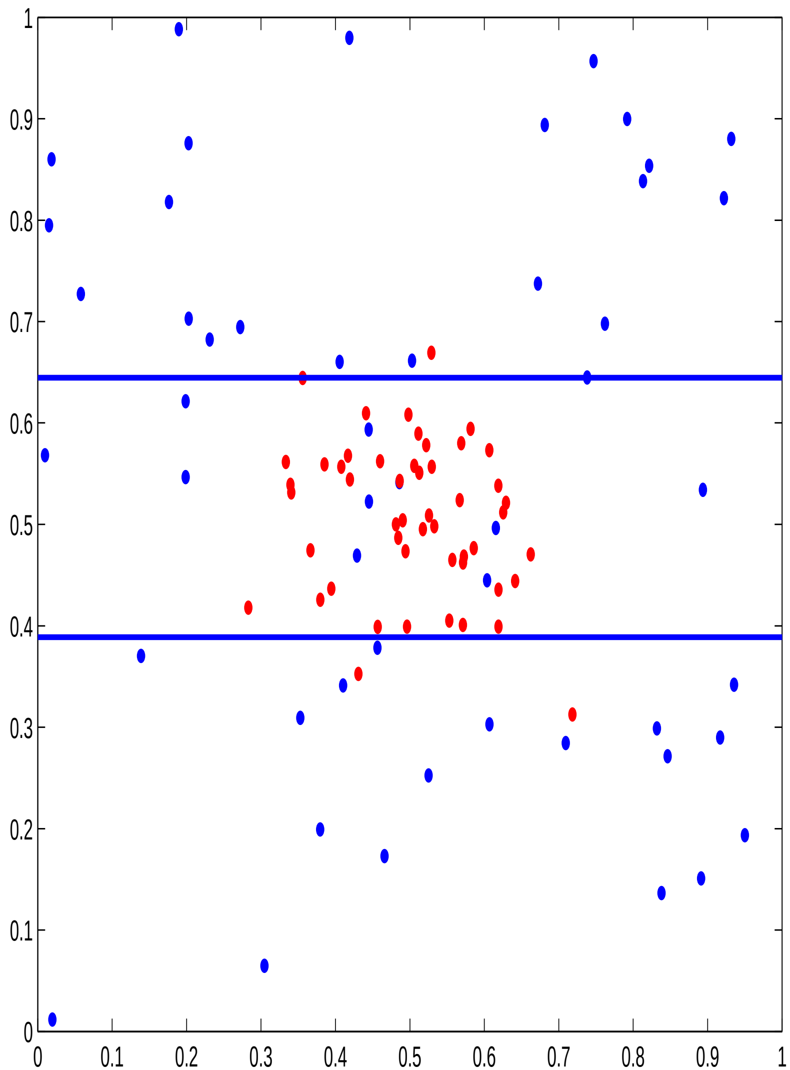
- Classification trees work well with high dimensional, mixed variable data
- Sequential greedy partition (CART algorithm)
  - Gini diversity index
- Partitions region into hyper-rectangles
  - Restriction on shape of B
- Allows weighting
  - If $\theta_{H_o} \neq 1$
- Grow multiple trees with randomness in splits
  - To overcome initial greedy splits

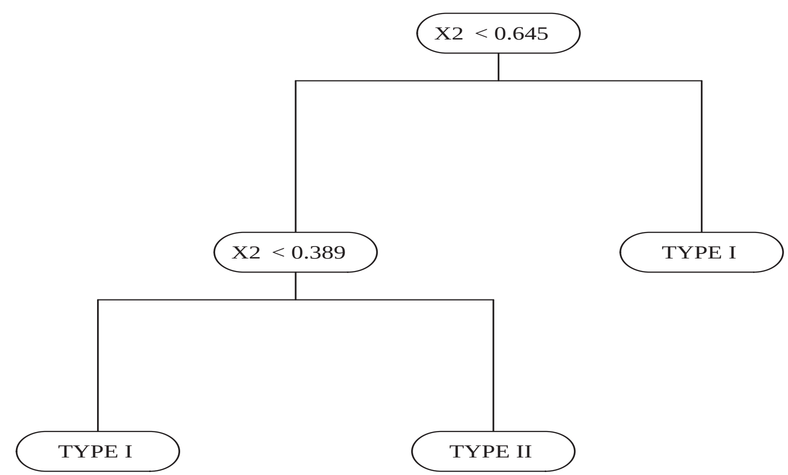
Classification Tree Example



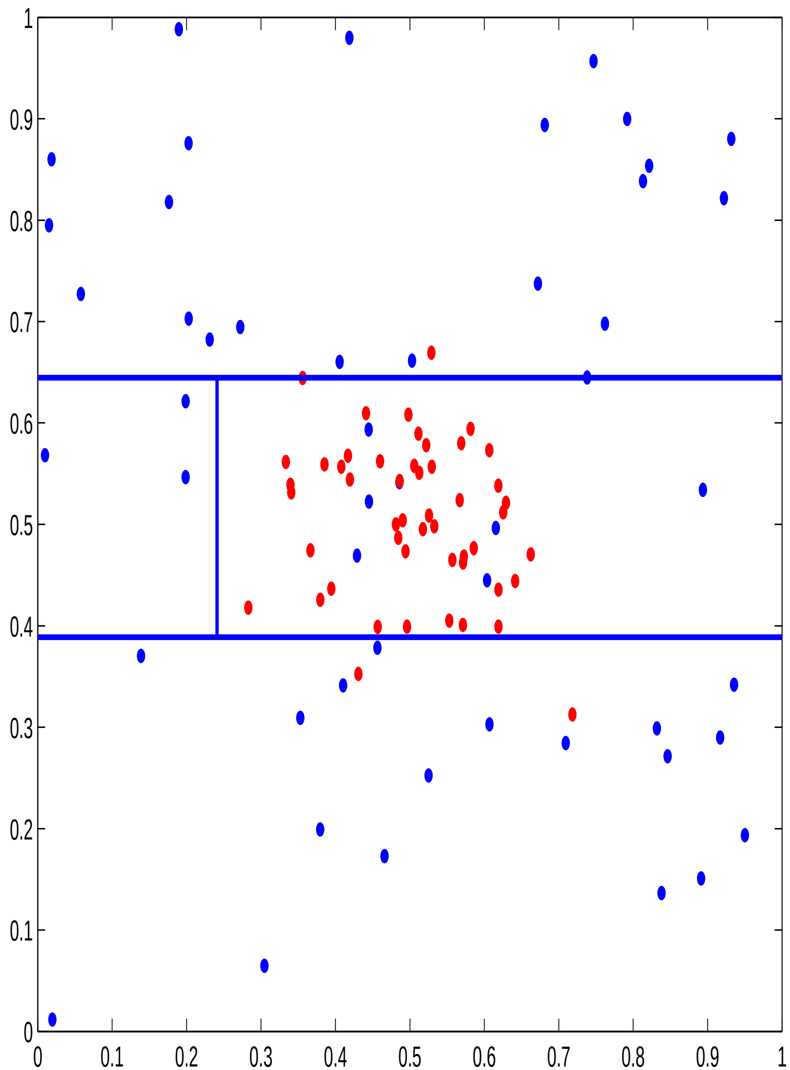


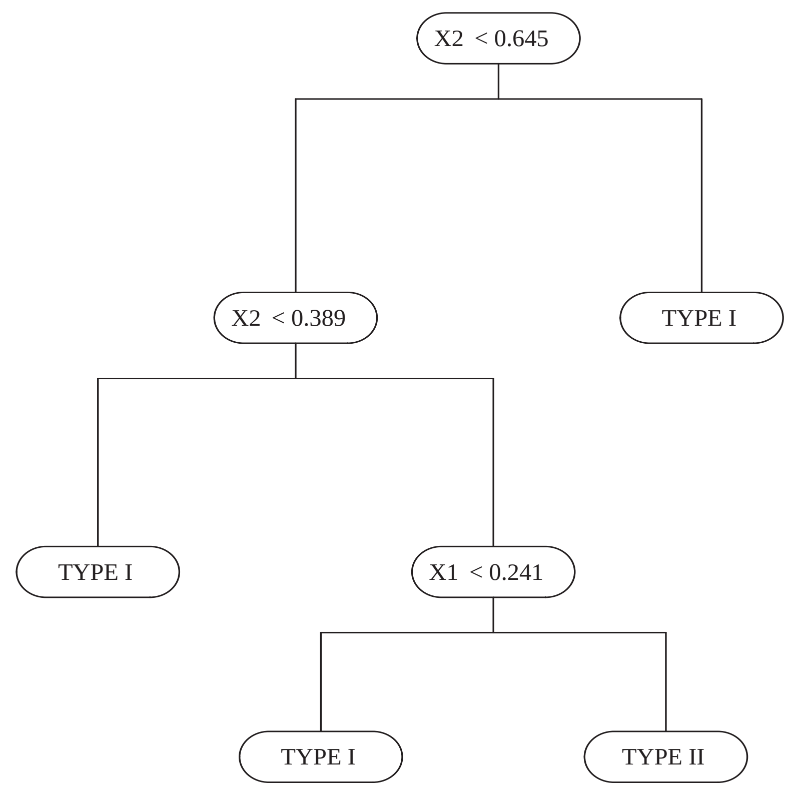
Classification Tree Example (2)





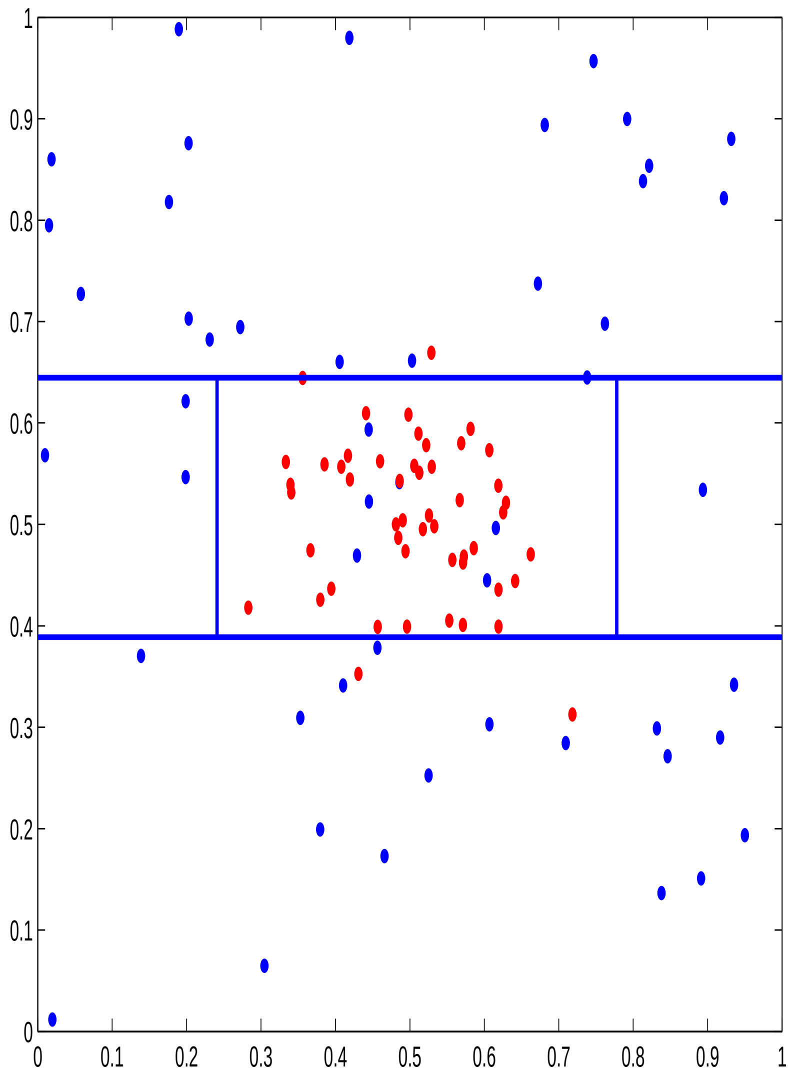
Classification Tree Example (3)

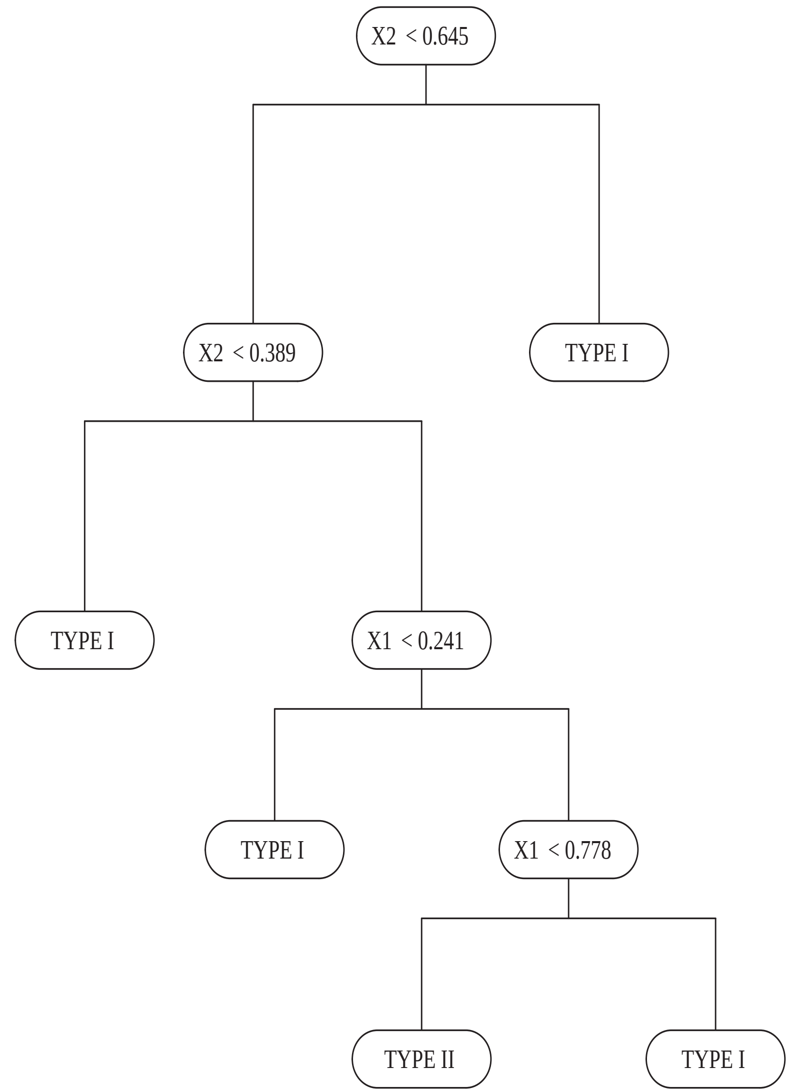




UNIVERSITY of VIRGINIA

Classification Tree Example (4)





UNIVERSITY of VIRGINIA

Significance - Monte Carlo

P-value

$$p_{obs} = Pr(T(\Omega) \leq T(\Omega^{obs}); H_o)$$

Reject H_o if $p_{obs} \leq \alpha$.

Monte Carlo

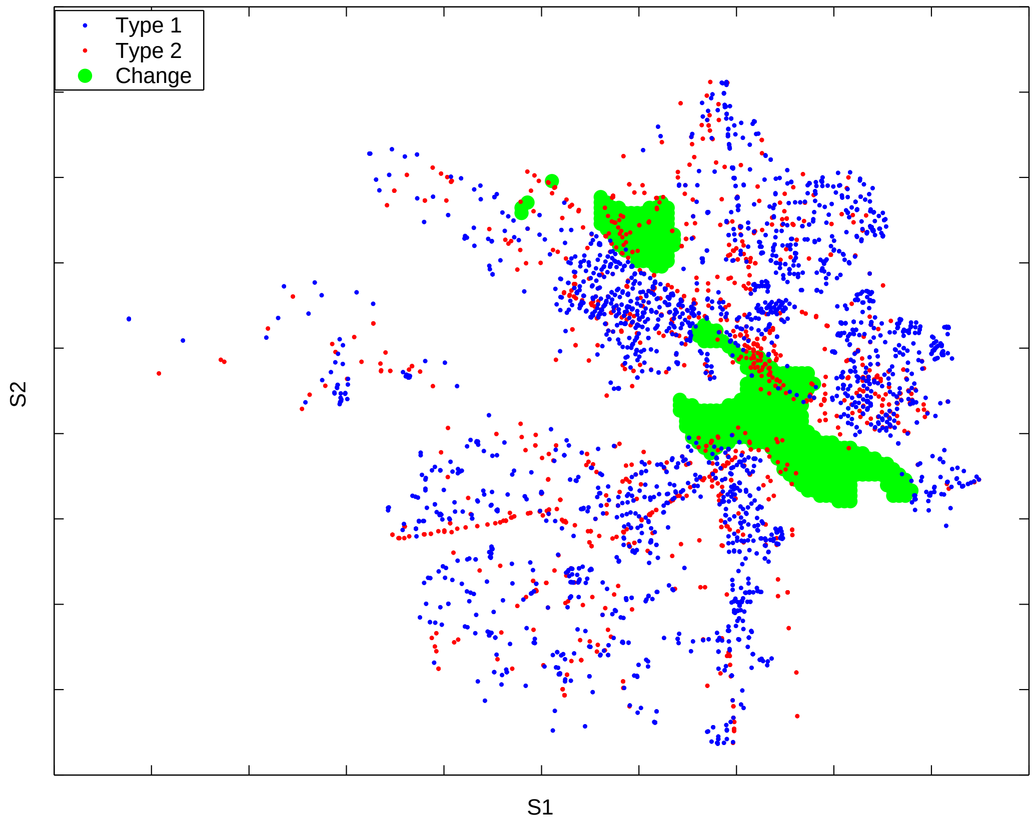
Generate $\{T(\Omega^{(m)}; H_o)\}_{m=1}^{M-1}$ by *random labeling* the event types according to:

$$f(k=1) = \theta_{H_o}(\theta_{H_o} + 1)^{-1}$$

Order these observations, $T(\Omega^{(m)})$, from smallest to largest and let l_m be the order of the m^{th} observation. The estimated p-value then becomes: $\hat{p}_{obs} = l_{obs}/M$

Example - Res vs. NonRes Density

B&E Crimes Richmond, VA 1997



Residential vs. Nonresidential B+E Density

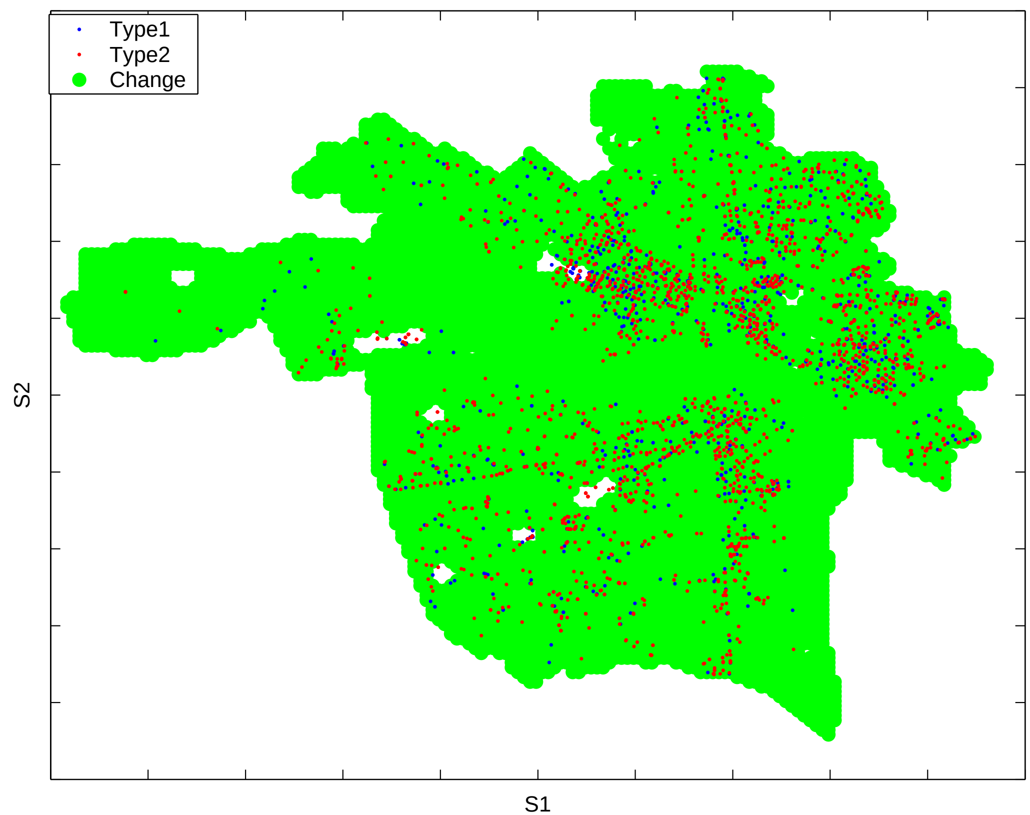
Residential = 2438 Events

Non-residential = 1171 Events

UNIVERSITY of VIRGINIA

Example - Hot vs. Cold Intensity

$$f(k=1 \mid \mathbf{s}) = \frac{|T_{hot}|}{|T_{cold}|} \cdot f(k=2 \mid \mathbf{s}) \Rightarrow \theta(\mathbf{s}) = \frac{107}{258} = .415$$



Hot vs. Cold B+E Intensity

$\hat{p}_{obs} = 38/100$, Don't reject H_o

Discussion

- Looking beyond spatial analysis alone
- High Dimensional
- Mixed Variables
- Similarities to Scan Stat